

SEGMENTING INFORMATION RECORDS WITH MISSING VALUES USING MULTIPLE PARTITION TREES

ABSTRACT OF THE INVENTION

- 5 A method and system for predicting the class membership of a record where information for one or more variables in the record is missing. Multiple classification trees are generated. A first classification tree is computed using a substantially complete set of information for all of the variables. Other classification trees are computed for different subsets of the
- 10 variables. Variables are selected for inclusion in a subset based on how strongly they influence the prediction of class membership. The first classification tree (based on the substantially complete set of information) is applied to a record with missing information. If missing information is needed by this tree in order to classify the record, another classification tree
- 15 that is not based on the missing variable is selected. The class membership for a record with information missing is predicted more accurately without substantially increasing the complexity of the prediction.